

PATENT

MS131753.03/MSFTP222USB

CERTIFICATE OF FACSIMILE TRANSMISSION

I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being submitted *via* the USPTO EFS Filing System on the date shown below to Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Date: October 15, 2007

/Michelle Pesek/

Michelle Pesek

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re patent application of:

Applicant(s): David E. Heckerman, *et al.*

Examiner: Wilbert L. Starks

Serial No: 09/430,767

Art Unit: 2129

Filing Date: October 29, 1999

Title: CLUSTER-BASED AND RULE BASED APPROACH FOR AUTOMATED  
WEB-BASED TARGETED ADVERTISING WITH QUOTAS

**Mail Stop Appeal Brief – Patents  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450**

---

**APPEAL BRIEF**

---

Dear Sir:

Appellant's representative submits this brief in connection with an appeal of the above-identified patent application. If any additional fees are due and/or are not covered by the credit card, the Commissioner is authorized to charge such fees to Deposit Account No. 50-1063 [MSFTP222USB].

**I. Real Party in Interest (37 C.F.R. §41.37(c)(1)(i))**

The real party in interest in the present appeal is Microsoft Corporation, the assignee of the subject application.

**II. Related Appeals and Interferences (37 C.F.R. §41.37(c)(1)(ii))**

Appellant, appellant's legal representative, and/or the assignee of the subject application are not aware of any appeals or interferences which may be related to, will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

**III. Status of Claims (37 C.F.R. §41.37(c)(1)(iii))**

Claims 1, 2, 4-36, 38 and 41-75 stand rejected by the Examiner. The rejection of claims 1, 2, 4-36, 38 and 41-75 is being appealed.

**IV. Status of Amendments (37 C.F.R. §41.37(c)(1)(iv))**

No claim amendments have been entered after the Final Office Action.

**V. Summary of Claimed Subject Matter (37 C.F.R. §41.37(c)(1)(v))****A. Independent Claim 1**

Independent claim 1 relates to a computer-implemented method. The method includes allocating each of a plurality of ads to at least one of a plurality of clusters based on a predetermined criterion accounting for at least a quota for each ad and a constraint for each cluster. (*See e.g.*, pg. 24, lns. 6-8; Figs. 2 and 3, elements 203, 205, 207, 213, 215, and 217; and Fig. 5.) The method also includes selecting an ad for a current cluster from ads allocated to the current cluster and effecting the ad. (*See e.g.*, pg. 11, lns. 18-20; pg. 25, lns. 3-7; and Fig. 5.)

**B. Independent Claim 36**

Independent claim 36 relates to a computer-implemented method. The method includes defining a plurality of clusters; each cluster corresponding to a group of users

who are most receptive to a given type of ad, defining the plurality of clusters comprises utilizing one of: user information obtained without monitoring; a Bayesian network; or a naïve-Bayes-network clustering approach. (See e.g., Figs. 2 and 3, elements 203, 205, 207, 213, 215, and 217; pg. 24, lns. 1-5; pg. 19, lns. 13-20; Fig. 4, and Fig. 5.) The method also includes allocating an ad having a particular type to at least one cluster based on the particular type of the ad and based on a predetermined criterion. (See e.g., pg. 24, lns. 6-8; and Fig. 5.)

**C. Independent Claim 46**

Independent claim 46 relates to a computer-implemented method. The method includes determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ . The method also includes outputting the allocation of each ad to at least one of the plurality of clusters. (See e.g., pg. 15, lns 13 to pg. 16, ln. 2, and Fig. 5.)

**D. Independent Claim 50**

Independent claim 50 relates to a computerized system. The system includes a database (606) storing a plurality of ads, each ad having a quota and an allocator (610) to allocate each of the plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least the quota for each ad and a constraint for each cluster. The system also includes a communicator (608) to select an ad for a current cluster from ads allocated to the current cluster and output the ad to a user. (See e.g., pg. 25, ln. 17 to pg. 26, ln. 19 and Fig. 6.)

**E. Independent Claim 53**

Independent claim 53 relates to a machine-readable medium having instructions stored thereon for execution by a processor to perform a method. The method includes allocating each of a plurality of ads to at least one of a plurality of clusters based on a predetermined criterion accounting for at least a quota for each ad and a constraint for

each cluster. (*See e.g.*, pg. 24, lns. 6-8; Figs. 2 and 3, elements 203, 205, 207, 213, 215, and 217; and Fig. 5.) The method also includes selecting an ad for a current cluster from ads allocated to the current cluster and displaying the ad. (*See e.g.*, pg. 11, lns. 18-20; pg. 25, lns. 3-7; and Fig. 5.)

**F. Independent Claim 59**

Independent claim 59 relates to a machine-readable medium having instructions stored thereon for execution by a processor to perform a method. The method includes determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ . The method also includes outputting the allocation of each ad to at least one of the plurality of clusters. (*See e.g.*, pg. 15, lns 13 to pg. 16, ln. 2, and Fig. 5.)

**G. Independent Claim 62**

Independent claim 62 relates to a computer-implemented method. The method includes applying each of at least one first ad to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second ads, to determine a second ad for each of the at least one first ad. The method also includes effecting the second ad for each of the at least one first ad. (*See e.g.*, pg. 27, ln. 23 to pg. 28, ln. 14; and Fig. 7.)

**H. Independent Claim 69**

Independent claim 69 relates to a computer-implemented method. The method includes determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activation of ads from training data and determining an ad and at least one binary feature providing a largest activation, each rule accounting for at least a quota for the ad. (*See e.g.*, pg. 29, lns. 18-20; and Fig. 8.) The method also includes generating a rule based on the ad and the at least one binary feature providing the largest activation, removing records from the training data matching the

rule generated, and repeating to generate another, lower-ordered rule while at least one significant correlation still exists. (*See e.g.*, pg. 30, lns. 4-22; and Fig. 8.)

**I. Independent Claim 71**

Independent claim 71 relates to a machine-readable medium having instructions stored thereon for execution by a processor to perform a method. The method includes applying each of at least one first ad to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second ads, to determine a second ad for each of the at least one first ad. (*See e.g.*, pg. 27, ln. 23 to pg. 28, ln. 14; and Fig. 7.) The method also includes effecting the second ad for each of the at least one first ad. (*See e.g.*, pg. 11, ln. 20; and Fig. 5.)

**J. Independent Claim 75**

Independent claim 75 relates to a machine-readable medium having instructions stored thereon for execution by a processor to perform a method. The method includes determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of ad from training data and determining an ad and at least one binary feature providing a largest activation, each rule accounting for at least a quota for the ad. (*See e.g.*, pg. 29, lns. 18-20; and Fig. 8.) The method further includes generating a rule based on the ad and the at least one binary feature providing the largest activation, removing records from the training data matching the rule generated, and repeating to generate another, lower-ordered rule while at least one significant correlation still exists. (*See e.g.*, pg. 30, lns. 4-22; and Fig. 8.)

**VI. Grounds of Rejection to be Reviewed (37 C.F.R. §41.37(c)(1)(vi))**

**A.** Whether claims 1, 2, 4-36, 38, and 41-76 are unpatentable under 35 U.S.C. §101 as being directed to non-statutory subject matter.

**B.** Whether claims 1, 2, 4-36, 38, and 41-76 are unpatentable under 35 U.S.C. §112, first paragraph because current case law and the MPEP require such a rejection if a

§101 rejection is given).

C. Whether claims 1 and 2 are unpatentable under 35 U.S.C. §102(e) as being anticipated by Ballard (U.S. 6,182,050).

**VII. Argument (37 C.F.R. §41.37(c)(1)(vii))**

**A. Rejection of Claims 1, 2, 4-36, 38, and 41-76 Under 35 U.S.C. §101**

Claims 1, 2, 4-36, 38, and 41-76 stand rejected under 35 U.S.C. §101 as being directed to non-statutory subject matter. This rejection is improper because the subject claims recite statutory subject matter as defined by 35 U.S.C. § 101 and, further, the subject claims relate to practical applications.

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

and

Because the claimed process applies the Boolean principle *to produce a useful, concrete, tangible result* ... on its face the claimed process comfortably falls within the scope of §101. *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1358 (Fed.Cir. 1999). The inquiry into patentability requires an examination of the contested claims to see if the claimed subject matter, as a whole, is a disembodied mathematical concept representing nothing more than a "law of nature" or an "abstract idea," or if the mathematical concept has been *reduced to some practical application rendering it "useful."* *AT&T* at 1357 citing *In re Alappat*, 33 F.3d 1526, 31 1544, 31 U.S.P.Q.2D (BNA) 1545, 1557 (Fed. Cir. 1994).

The subject application relates to cluster-based and rule based approaches for targeted advertising with quotas. Ads, such as banner ads, house ads, and targeted advertising are used to advertise or bring awareness of items, services, events and so on to the consuming public. (See e.g., pg. 1, lns. 11-15; pg. 2, lns. 4-6; pg. 11, lns. 5-14; and

pg. 17, Ins. 5-9). In the Final Office Action, it was stated “Examiner is not of the opinion that a pure ‘advertisement’ is not ‘useful, concrete and tangible’ ... and is not a ‘substance’.” Based on a dictionary definition of “advertisement” it is asserted that:

“A mere ‘notice’ is not tangible. It is not a ‘substance’ and it need not involve any commerce. It is mere data display. As such, it is abstract and pure manipulations of ‘ads’ are likewise, abstract ideas.”

However, the purpose or *usefulness* of an ad is to draw the public’s attention to the item being advertised and such advertisements are more than “mere data display”. The decision whether to act on the ad (*e.g.*, involve commerce) is entirely left to the consuming public. Each individual can become aware of the item being advertised by the advertisement, which is tangible, useful and concrete (*e.g.*, perceived and drawing consuming public’s attention to something).

Even if “ads”, assuming *arguendo*, are abstract ideas, the subject claims go beyond mere manipulation of abstract items and a person of ordinary skill in the relevant art would appreciate the usefulness and practical application of the subject claims. Specific and substantial utilities of the subject claims relate to maximizing a click through rate for ads given a quota, allowing web site operators to maximize earning potential from advertising, adhering to various obligations relating to advertising within certain clusters (*e.g.*, adult-oriented ad might not be desired to be shown in a children-oriented cluster), as well as other utilities. (*See e.g.*, pg. 3, ln. 13 to pg. 4, ln. 4). In addition, some advertisers can be favored over other advertisers for various reasons. (*See e.g.*, pg. 16, Ins. 18-20). The ad can be effected, which can include displaying the ad or displaying a button on a web site for immediate purchase of an item. (*See e.g.*, pg. 11, ln. 18 to pg. 12, ln. 3 and pg. 28, Ins. 10-14). Effecting the ad can also relate to maximizing a possibility that a user (*e.g.*, consumer) will purchase something, “not just click on the ad.” (*See e.g.*, pg. 18, Ins. 16-17). Thus, the subject claims relate to practical applications and produce a useful, tangible and concrete result for both an advertiser and a web site operator in order to maximize earning potentials from advertising. (*See e.g.*, pg. 3, Ins. 17-19).

Therefore, based on the above, the subject claims recite a practical application, go beyond mere manipulation of abstract ideas and produce a concrete, tangible and useful result. Therefore, it is requested that this rejection be withdrawn and the subject claims allowed.

**B. Rejection of Claims 1, 2, 4-36, 38, and 41-76 Under 35 U.S.C. §112**

Claims 1, 2, 4-36, 38, and 41-76 stand rejected under 35 U.S.C. §112, first paragraph because current case law and the MPEP require such a rejection if a §101 rejection is given. This rejection should be withdrawn for at least the following reasons because, as discussed above, the subject claims disclose a practical application that produces a concrete, tangible and useful result and go beyond manipulation of abstract ideas. As such, the subject claims recite statutory subject matter as defined by 35 U.S.C. § 101. Therefore, this rejection should be withdrawn.

**C. Rejection of Claims 1 and 2 Under 35 U.S.C. §102(e)**

Claims 1 and 2 stand rejected under 35 U.S.C. §102(e) as being anticipated by Ballard (U.S. 6,182,050). This rejection should be withdrawn for at least the following reason. Ballard does not teach or suggest all limitations recited in the subject claims.

For a prior art reference to anticipate, 35 U.S.C. §102 requires that “each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference.” *In re Robertson*, 169 F.3d 743, 745, 49 USPQ2d 1949, 1950 (Fed. Cir. 1999) (*quoting Verdegaal Bros., Inc. v. Union Oil Co.*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987)).

“To establish inherency, the extrinsic evidence ‘must make clear that the missing descriptive matter is necessarily present in the thing described in the reference, and that it would be so recognized by persons of ordinary skill.’” *Id.* (*quoting Continental Can co. v. Monsanto Co.*, 948 F.2d 1264, 1268, 20 USPQ2d 1746, 1749 (Fed. Cir. 1991)). “Inherency, however, may not be established by probabilities or possibilities. The mere fact that a certain thing may result from a given set of circumstances is not sufficient.”



*Mehl/Biophile Int'l Corp. v. Milgraum*, 192 F.3d 1362, 1365, 52 USPQ2d 1303, 1305 (Fed. Cir. 1999), reh'g denied, 1999 U.S. App. LEXIS 31386 (Fed. Cir. Oct. 27, 1999) (quoting *In re Oelrich*, 666 F.2d 578, 581, 212 USPQ 323, 326 (CCPA 1981)).

Independent claim 1, recites *a computer-implemented method comprising allocating each of a plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least a quota for each ad and a constraint for each cluster, selecting an ad for a current cluster from ads allocated to the current cluster and effecting the ad*. For example, the quota for each ad and the constraint for each cluster can be used to maximize a number of click through for all the ads, given the quotas and constraints. (See e.g., pg. 15, lns. 8-11). Ballard does not expressly nor inherently describe such novel features.

Instead, Ballard relates to advertisement distribution *based on demographic data*. (See e.g., col. 9, lns. 39-48). These demographics are characteristics of human populations and population segments and can include hobbies, interest, credit history, travel history and past purchasing history. (See e.g., col. 1, lns. 21-33 and col. 7, lns. 3-13). However, Ballard fails to teach or even suggest accounting for at least a quota for each ad and a constraint for each cluster, as claimed.

Since Ballard does not expressly or inherently describe all limitations recited in the subject claims this rejection should be withdrawn and the subject claims allowed.

**D. Conclusion**

For at least the above reasons, the claims currently under consideration are believed to be patentable over the cited references. Accordingly, it is respectfully requested that the rejections of claims 1, 2, 4-36, 38 and 41-75 be reversed.

If any additional fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063[MSFTP222USB].

Respectfully submitted,  
AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/

Himanshu S. Amin  
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP  
24<sup>th</sup> Floor, National City Center  
1900 East 9<sup>th</sup> Street  
Telephone: (216) 696-8730  
Facsimile: (216) 696-8731

**VIII. Claims Appendix (37 C.F.R. §41.37(c)(1)(viii))**

1. A computer-implemented method comprising:  
 allocating each of a plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least a quota for each ad and a constraint for each cluster;  
 selecting an ad for a current cluster from ads allocated to the current cluster; and,  
 effecting the ad.
2. The method of claim 1, effecting the ad comprises displaying the ad.
3. (Cancelled).
4. The method of claim 2, the predetermined criterion further accounts for a particular one of the plurality of ads restricted from being shown in a particular one or more of the plurality of clusters.
5. The method of claim 2, the predetermined criterion comprises maximizing an expression  $\sum_j p_{ij} x_{ij}$ , where  $p_{ij}$  comprises a probability that a user in cluster  $j$  will actuate ad  $i$ .
6. The method of claim 5, the predetermined criterion further comprises maximizing the expression subject to a constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .
7. The method of claim 5, the predetermined criterion further comprises maximizing the expression subject to a constraint  $\sum_i x_{ij} = c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

8. The method of claim 5, the predetermined criterion comprises maximizing the expression subject to a first constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ , and a second constraint  $\sum_i x_{ij} = c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ , such that the expression, the first constraint and the second constraint define a linear program.
9. The method of claim 8, the linear program is solved by the Simplex Algorithm.
10. The method of claim 2, allocating each of a plurality of ads to at least one of the plurality of clusters comprises determining for each ad in each cluster a probability that a user in the cluster will actuate the ad.
11. The method of claim 10, the probability that a user in the cluster will actuate the ad comprises the probability that a user in the cluster will click on the ad.
12. The method of claim 10, determining for each ad in each cluster a probability that a user in the cluster will actuate the ad comprises inputting training data from which to determine for each ad in each cluster the probability that a user in the cluster will actuate the ad.
13. The method of claim 10, determining for each ad in each cluster a probability that a user in the cluster will actuate the ad comprises utilizing at least one of: a maximum likelihood approach, a MAP method approach, and, a hierarchical Bayesian approach.
14. The method of claim 2, the predetermined criterion comprises maximizing an expected number of actuations of the plurality of ads, given the quota for each ad and the constraint for each cluster.

15. The method of claim 2, the constraint for each cluster comprises a total number of times the cluster is visited by any user.
16. The method of claim 2, the quota for each ad comprises a total number of times that the ad must be displayed.
17. The method of claim 2, the criterion comprises favoring at least one ad over other ads within the plurality of ads in allocating the at least one ad.
18. The method of claim 2, the criterion comprises accounting for at least one house ad.
19. The method of claim 2, the predetermined criterion comprises minimizing an expression  $\sum_{ij} p_{ij} x_{ij}$ , where  $p_{ij}$  comprises a probability that a user in cluster  $j$  will actuate ad  $i$ .
20. The method of claim 2, the predetermined criterion comprises maximizing an expression  $\sum_{ij} \alpha_i p_{ij} x_{ij}$ , where  $p_{ij}$  comprises a probability that a user in cluster  $j$  will actuate ad  $i$ , and  $\alpha_i$  comprises a coefficient for the ad  $i$  to indicate weighting of the ad  $i$ .
21. The method of claim 5, the predetermined criterion further comprises maximizing the expression subject to a constraint  $x_{ij} = 0$  for a particular ad  $i$  within a particular cluster  $j$ , where  $x_{ij}$  comprises a total number of times the ad  $i$  is shown in the cluster  $j$ .
22. The method of claim 5, the predetermined criterion further comprises maximizing the expression subject to a constraint  $\sum_i x_{ij} \leq c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

23. The method of claim 10, the probability that a user in the cluster will actuate the ad comprises the probability that a user in the cluster will make a purchase based on the ad.
24. The method of claim 2, the method includes first initially defining the plurality of clusters.
25. The method of claim 24, defining the plurality of clusters comprises utilizing user information obtained without monitoring.
26. The method of claim 24, utilizing user information obtained without monitoring comprises utilizing a category tag (e.g., page group) of the page on which the ad is to be displayed.
27. The method of claim 25, utilizing user information obtained without monitoring comprises utilizing user information obtained from the user via a questionnaire.
28. The method of claim 24, defining the plurality of clusters comprises utilizing a preexisting plurality of groups as the plurality of clusters.
29. The method of claim 24, defining the plurality of clusters comprises utilizing a Bayesian network.
30. The method of claim 24, defining the plurality of clusters comprises utilizing a naïve-Bayes-network clustering approach.
31. The method of claim 30, utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture.
32. The method of claim 30, utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture recursively to construct a hierarchy of

clusters.

33. The method of claim 30, utilizing a Bayesian network clustering approach comprises training a Bayesian network using a stochastic gradient descent technique.

34. The method of claim 30, utilizing a Bayesian network clustering approach comprises employing a single hidden variable having a plurality of values.

35. The method of claim 30, utilizing a Bayesian network clustering approach comprises employing a plurality of hidden variables, each having two values.

36. A computer-implemented method comprising:

defining a plurality of clusters, each cluster corresponding to a group of users who are most receptive to a given type of ad, defining the plurality of clusters comprises utilizing one of:

user information obtained without monitoring;

a Bayesian network; or

a naïve-Bayes-network clustering approach; and,

allocating an ad having a particular type to at least one cluster based on the particular type of the ad and based on a predetermined criterion.

37. (Cancelled).

38. The method of claim 36, utilizing user information obtained without monitoring comprises utilizing user information obtained from the user via a questionnaire.

39. (Cancelled).

40. (Cancelled).

41. The method of claim 36, utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture.
42. The method of claim 36, utilizing a Bayesian network clustering approach comprises utilizing a hierarchical bottleneck architecture.
43. The method of claim 36, utilizing a Bayesian network clustering approach comprises training a Bayesian network using a stochastic gradient descent technique.
44. The method of claim 36, utilizing a Bayesian network clustering approach comprises employing a single hidden variable having a plurality of values.
45. The method of claim 36, utilizing a Bayesian network clustering approach comprises employing a plurality of hidden variables, each having two values.
46. A computer-implemented method comprising:
  - determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ ; and,
  - outputting the allocation of each ad to at least one of the plurality of clusters.
47. The method of claim 46, determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises maximizing an expression  $\sum_{ij} p_{ij} x_{ij}$ , where  $p_{ij}$  comprises a probability that a user in cluster  $j$  will actuate ad  $i$ , given the constraint.
48. The method of claim 46, determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises determining the allocation for each of the plurality of ads to at least one of the plurality of clusters further given a constraint



$\sum_i x_{ij} = c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

49. The method of claim 46, further comprising:  
selecting an ad for a current cluster from the allocation of each ad to the current cluster; and,  
displaying the ad.
50. A computerized system comprising:  
a database storing a plurality of ads, each ad having a quota;  
an allocator to allocate each of the plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least the quota for each ad and a constraint for each cluster; and,  
a communicator to select an ad for a current cluster from ads allocated to the current cluster and output the ad to a user.
51. The system of claim 50, at least one of the allocator and the communicator comprises a computer program executed from a computer-readable medium by a processor.
52. The system of claim 50, the database is stored as data on a computer-readable medium.
53. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:  
allocating each of a plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least a quota for each ad and a constraint for each cluster;  
selecting an ad for a current cluster from ads allocated to the current cluster; and,  
displaying the ad.

54. The medium of claim 53, the predetermined criterion comprises maximizing an expression  $\sum_j p_j x_j$ , where  $p_j$  comprises a probability that a user in cluster  $j$  will actuate ad  $I$

55. The medium of claim 54, the predetermined criterion further comprises maximizing the expression subject to a constraint  $\sum_j x_j = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_j$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

56. The medium of claim 54, the predetermined criterion further comprises maximizing the expression subject to a constraint  $\sum_i x_{ij} = c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

57. (The medium of claim 53, allocating each of a plurality of ads to at least one of the plurality of clusters comprises determining for each ad in each cluster a probability that a user in the cluster will actuate the ad.

58. The medium of claim 53, the predetermined criterion comprises maximizing an expected number of actuations of the plurality of ads, given the quota for each ad and the constraint for each cluster.

59. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint  $\sum_j x_{ij} = q_i$ , where  $q_i$  comprises a quota for ad  $i$ , and  $x_{ij}$

comprises a total number of times ad  $i$  is shown in cluster  $j$ ; and,  
 outputting the allocation of each ad to at least one of the plurality of clusters.

60. The medium of claim 59, determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises maximizing an expression  $\sum_{ij} p_{ij} x_{ij}$ ,

where  $p_{ij}$  comprises a probability that a user in cluster  $j$  will actuate ad  $i$ , given the constraint.

61. The medium of claim 59, determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises determining the allocation for each of the plurality of ads to at least one of the plurality of clusters further given a constraint  $\sum_i x_{ij} = c_j$ , where  $c_j$  comprises a constraint for cluster  $j$ , and  $x_{ij}$  comprises a total number of times ad  $i$  is shown in cluster  $j$ .

62. A computer-implemented method comprising:  
 applying each of at least one first ad to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second ads, to determine a second ad for each of the at least one first ad; and,  
 effecting the second ad for each of the at least one first ad.

63. The method of claim 62, each first ad comprises at least information about a user, and a web page currently being browsed by the user.

64. The method of claim 62, effecting the second ad comprises displaying the ad.

65. The method of claim 62, further comprising generating the ordered set of rules based on training data.

66. The method of claim 65, generating the ordered set of rules comprises:  
determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of second ads of the training data;  
determining a second ad and at least one binary feature providing a largest activation; and,  
generating a rule based on the second ad and the at least one binary feature providing the largest activation.
67. The method of claim 66, generating the ordered set of rules further comprises:  
removing records from the training data matching the rule generated; and,  
repeating to generate another, lower-ordered rule while at least one significant correlation still exists.
68. The method of claim 66, determining at least one significant correlation comprises utilizing one of: Chi-squared method, Fisher exact test method, and Bayesian model selection method.
69. A computer-implemented method comprising:  
determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activation of ads from training data;  
determining an ad and at least one binary feature providing a largest activation, each rule accounting for at least a quota for the ad;  
generating a rule based on the ad and the at least one binary feature providing the largest activation;  
removing records from the training data matching the rule generated; and,  
repeating to generate another, lower-ordered rule while at least one significant correlation still exists.
70. The method of claim 69, each ad comprises an ad.

71. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

- applying each of at least one first ad to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second ads, to determine a second ad for each of the at least one first ad; and,
- effecting the second ad for each of the at least one first ad.

72. The medium of claim 71, the method further-comprising generating the ordered set of rules based on training data.

73. The medium of claim 71, each first ~~mad~~ ad comprises at least information about a user, and a web page currently being browsed by the user, and each second ad comprises an ad.

74. The medium of claim 71, generating the ordered set of rules comprises:

- determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of second ad of the training data;
- determining a second ad and at least one binary feature providing a largest activation;
- generating a rule based on the second ad and the at least one binary feature providing the largest activation;
- removing records from the training data matching the rule generated; and,
- repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

75. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

- determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of ad from training data;
- determining an ad and at least one binary feature providing a largest activation,

each rule accounting for at least a quota for the ad;

generating a rule based on the ad and the at least one binary feature providing the largest activation;

removing records from the training data matching the rule generated; and,

repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

76. (Cancelled)

**IX. Evidence Appendix (37 C.F.R. §41.37(c)(1)(ix))**

None.

**X. Related Proceedings Appendix (37 C.F.R. §41.37(c)(1)(x))**

None.